

Introduction to Bioinformatics

1. Databases and Literature Searches

Benjamin F. Matthews

United States Department of Agriculture

Soybean Genomics and Improvement

Laboratory

Beltsville, MD 20708

matthewb@ba.ars.usda.gov

Outline of course

- | | |
|--|-------------------------------------|
| ✖ Introduction to Bioinformatics | ✖ Phylogenetic trees |
| ✖ Databases | ✖ Restriction analysis |
| ✖ Literature searches | ✖ Primer design |
| ✖ DNA sequence retrieval | ✖ Genomic sequencing and annotation |
| ✖ DNA analysis <ul style="list-style-type: none">• Translation• alignment | ✖ Networks and pathways |
| ✖ Protein retrieval and analysis | ✖ Gene expression |

What we will cover today

- * What is bioinformatics
- * Databases
- * Literature searches

Bioinformatics

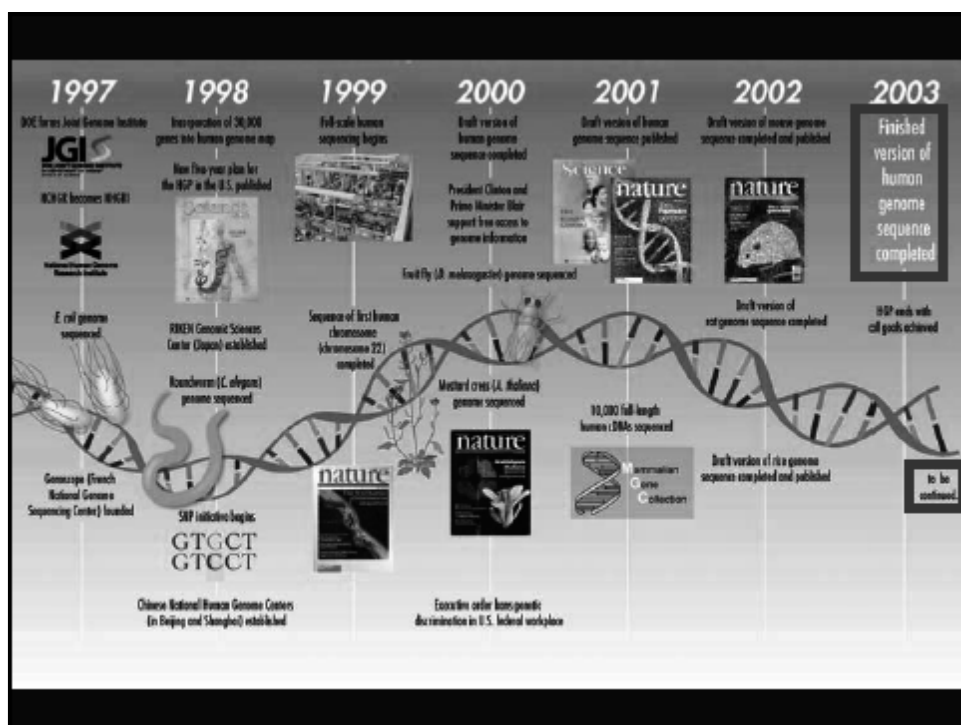
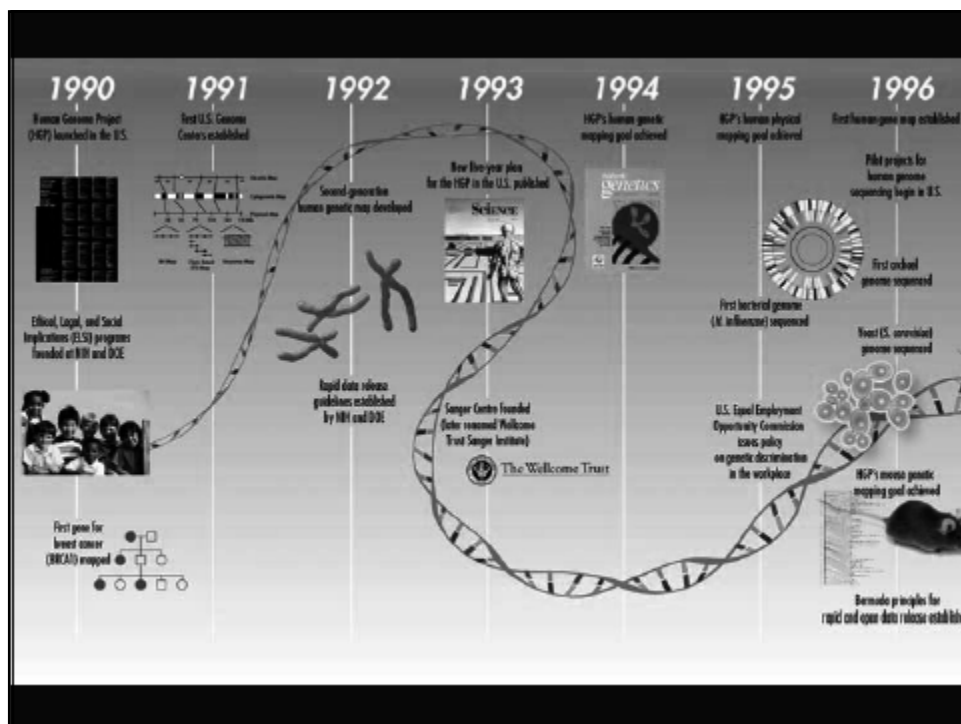
- the use of computers to collect, analyze, and interpret biological information
- a set of software tools and a database

Genomics Revolution

- * Costs of DNA sequencing and novel techniques (eg. microarrays) have fallen dramatically
- * Huge amount of genomics sequence is available
- * Novel techniques are more readily available
- * Huge amounts of data is moving biology toward becoming an information science

New challenge for biologists

- * Organize and mine data
- * Identify genes
- * Determine gene and protein function
- * Identify important and useful genes
 - ◆ Diagnostic
 - ◆ Modification
 - ◆ New value-added traits



So much data!

- * Need to store data in organized fashion
- * Need rapid retrieval of portions of data
- * Need to compare data
- * Need to annotate data

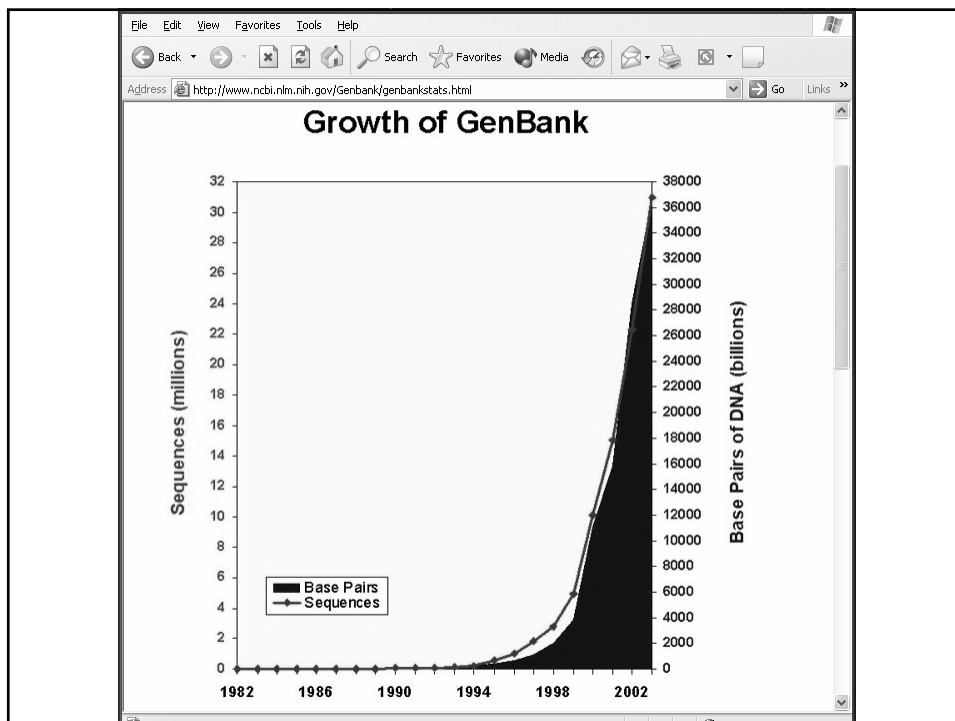
National Center for Biotechnology Information
<http://www.ncbi.nlm.nih.gov/>

Created as a part of NLM in 1988

- To establish public databases
 - U.S. National DNA Sequence Database
- To perform research in computational biology
- To develop software tools for sequence analysis
- To disseminate biomedical information

GenBank

- ✧ <http://www.ncbi.nlm.nih.gov/Genbank/index.html>
- ✧ Nucleotide sequences
- ✧ >130,000 organisms
- ✧ Annotated records with coding region features and amino acid translations



Genome update

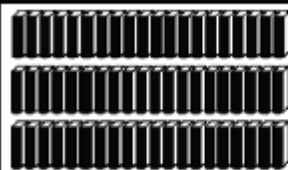
✱ Numerous genome drafts

- ◆ 155 bacteria
- ◆ 18 archaea
- ◆ numerous viruses
- ◆ Fungi (12 or more)
- ◆ Animal
 - *Caenorhabditis elegans*, *Drosophila*, mosquito, honey bee, zebra fish, chicken, mouse, rat, dog, chimpanzee, human
- ◆ Plant
 - *Arabidopsis*, Rice

✱ Many more under construction

Genome Sizes

Human Genome
Mouse Genome



~3,000,000,000 bp

Fruit Fly Genome



~160,000,000 bp

Nematode Genome



~100,000,000 bp

Yeast Genome



~15,000,000 bp

***E. coli* Genome**

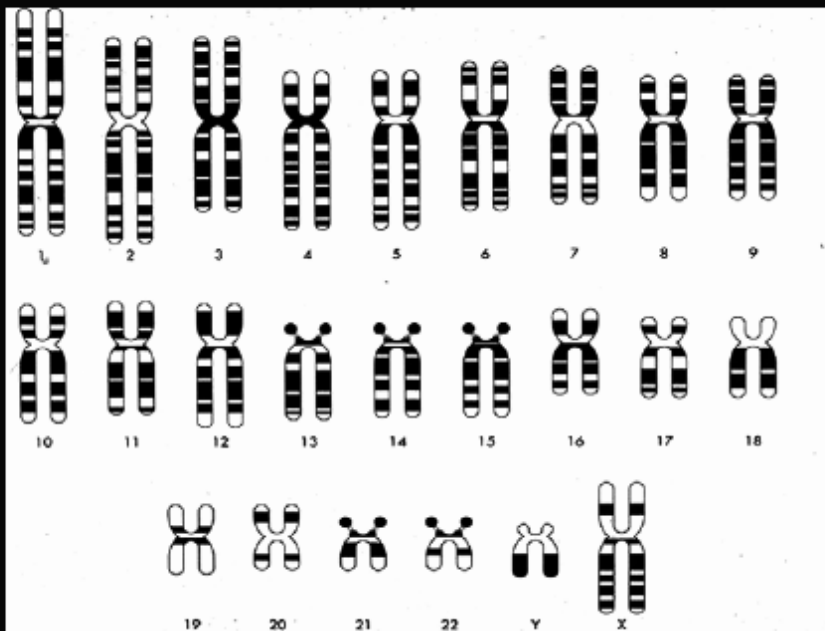


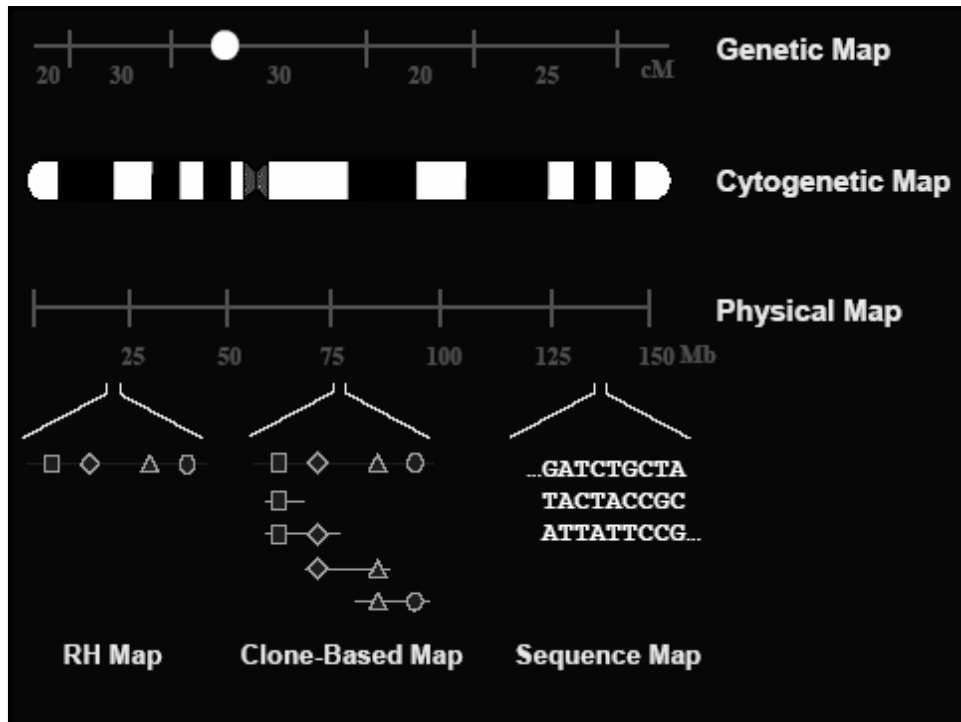
~5,000,000 bp

Human Genome

- * 32 billion base pairs
- * Close to a complete list of human genes and proteins
- * Challenge is to understand the genes, proteins and biological processes, so we can use this information to help the world

The Human Cytogenetic Map





Curriculum for bioinformatics students

* Biology

- ◆ Molecular biology, biochemistry, physiology

* Statistics, mathematics

- ◆ Data analysis, Algorithm development

* Computer programming

- ◆ Software design and development
- ◆ Perl, C++

* Database design and management

- ◆ SQL

* Web site design and management

* Use of current bioinformatics software packages and databases

A biology student

- * Strong biology background
- * Use of existing databases
- * Use of existing software

- * Does not necessarily need programming language

Databases

A Database

- ✧ Self-describing collection of integrated records
 - ✧ Computerized
 - ✧ Complete with language for access
-

- ✧ There are many databases and software tools available to mine the databases
- ✧ Provide permanent storage of information
- ✧ Application programs allow users to manipulate information

Biological Database

- ✧ A large, organized body of persistent data
- ✧ Associated with computerized software
 - ◆ Analyze
 - ◆ Update
 - ◆ Query
 - ◆ Retrieve
- ✧ Easy access to information
- ✧ Method of extraction of needed information only

Record

- * a representation of some physical or conceptual object
- * Eg. A record of business customers
 - ◆ A record contains **attributes**
 - ◆ Eg. Name, address, telephone number
- * An individual name, address, telephone number is **data**

Database Management System (DBMS)

- * A set of programs used to define, administer, and process databases

Conventional Operating Systems (Hierarchical)

- ✧ Information was placed in a record and stored in a file system
- ✧ Application programs were written to access information as needs arose
- ✧ As needs change, new permanent files must be created and new application programs must be written

Relational Database

- ✧ Can change database structure without rewriting applications
- ✧ Data may be retrieved from various files and from various formats
- ✧ Reduce redundancy and inconsistency
 - ◆ Reduces storage and access costs
- ✧ Easier access of information
 - ◆ New application programs need not be written
- ✧ Easier to enforce consistency constraints
- ✧ Easier to enforce security constraints
- ✧ Therefore it is more flexible

Some Biological Databases

✧ Literature (PubMed and Agricola)

- ◆ Journals
- ◆ books

✧ DNA (GenBank)

- ◆ cDNA and EST
- ◆ Whole genome

✧ Protein (Erez Protein and ExPASy)

- ◆ Sequence
- ◆ Structure

✧ Metabolic (KEGG regulatory pathways)

✧ Organism

- ◆ Human, mouse, drosophila, Arabidopsis

Major Molecular Databases

- NCBI (National Center for Biotechnology Information):
<http://www.ncbi.nih.gov>
 - Contains GenBank, PubMed, other databases, tools & links
- GenBank: <http://www.ncbi.nih.gov/Genbank/>
- EMBL (European Molecular Biology Laboratory)
<http://www.ebi.ac.uk/embl/>
- PDB (Protein Data Bank):
 - <http://www.rcsb.org/pdb/>

Major Molecular Databases

- PIR (Protein Information Resource):
 - ♦ <http://pir.georgetown.edu/>
- ExPasy proteomics server (Expert Protein Analysis System):
 - ♦ <http://us.expasy.org/>
- KEGG (Kyoto Encyclopedia of Genes and Genomes):
 - ♦ <http://www.genome.ad.jp/kegg/kegg2.html>

National Center for Biotechnology Information
<http://www.ncbi.nlm.nih.gov/>

Created as a part of NLM in 1988

- To establish public databases
 - U.S. National DNA Sequence Database
- To perform research in computational biology
- To develop software tools for sequence analysis
- To disseminate biomedical information

NCBI houses several databases

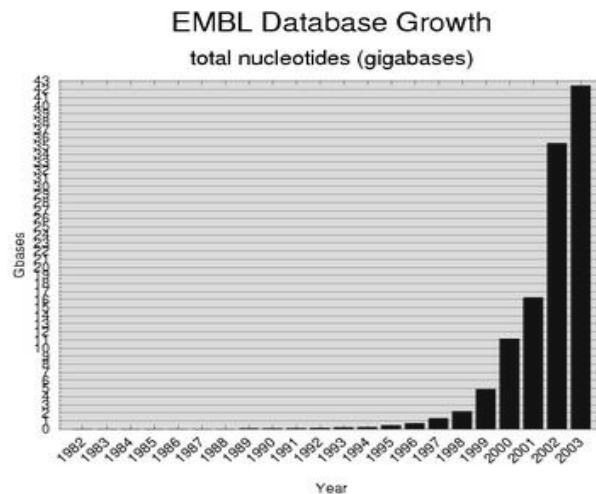
- ✖ Entrez
- ✖ PubMed
- ✖ Nucleotide
- ✖ Protein
- ✖ Genome
- ✖ Structure
- ✖ Taxonomy

<http://www.ncbi.nlm.nih.gov/>

EMBL Nucleotide Sequence Database

- ✖ Website: <http://www.ebi.ac.uk/embl/>
- ✖ Maintained at Europe Institute of Bioinformatics and is Europe's primary nucleotide sequence resource
- ✖ Include both DNA and RNA sequences
- ✖ From direct submissions of individual researchers, genome sequencing projects and patent applications
- ✖ Daily exchanges with GenBank and DDBJ

EMBL Database Statistics



ExPasy

- * Abbreviated from Expert Protein analysis System
- * Web site: <http://us.expasy.org/>
- * The ExPASy server is maintained by the Swiss Institute of Bioinformatics
- * Contains molecular databases, tools, and links

KEGG



- * GeneUniverse
 - ◆ Genes and gene products in complete genomes
- * Chemical networks
 - ◆ Chemical compounds
 - ◆ Chemical reactions
- * Protein Networks
 - ◆ Metabolic pathways
 - ◆ Regulatory pathways
 - ◆ Relationships

Databases

- * Organize data
- * Allows data submission and access
- * Resource for other databases and tools
- * Find what you want...fast

Database components

- ✖ Define and describe
- ✖ Unique identifier
- ✖ Update version
- ✖ Links to other databases
- ✖ Documentation
- ✖ Submission/update/correction process
- ✖ Retrieval system
- ✖ Standards
 - ◆ Universal if appropriate

Retrieval system

- ✖ User interface
 - ◆ Computer
- ✖ Batch mode
 - ◆ Avoid repetitive process
- ✖ Structured queries or SQL (Structured Query Language) access
 - ◆ Allows mining of data
 - ◆ User-friendly data mining
- ✖ Full-dump
 - ◆ Allows placement of data on other computers
- ✖ All data accessible

Retrieval System (cont.)

- ✧ Documentation

- ◆ How things work

- ✧ Link definitions

- ◆ What is at each spot

- ✧ User support

- ◆ If something goes wrong

Cost

- ✧ Hardware-computer, screen, keyboard, mouse, extra storage

- ✧ Software

- ✧ Production cost

- ◆ Training
 - ◆ data accumulation and entry

- ✧ Usage cost

- ◆ Computer, software, training, internet access

Literature databases and literature searches

You Define A Problem

✧ Target problem area

- ◆ Meet with farmers, extension agents, colleagues
- ◆ Read agricultural reports and literature
- ◆ Attend meetings

✧ Study problem

- ◆ Read literature
- ◆ Talk with colleagues
- ◆ Go to meetings and seminars

Define Project Goals

- ✧ Intensive, focused literature search
- ✧ Where can you conduct a literature search on-line?

Literature searches

- ✧ NCBI Pub Med
 - <http://www.ncbi.nlm.nih.gov/PubMed/>
- ✧ USDA Agricola
 - <http://agricola.nal.usda.gov/>

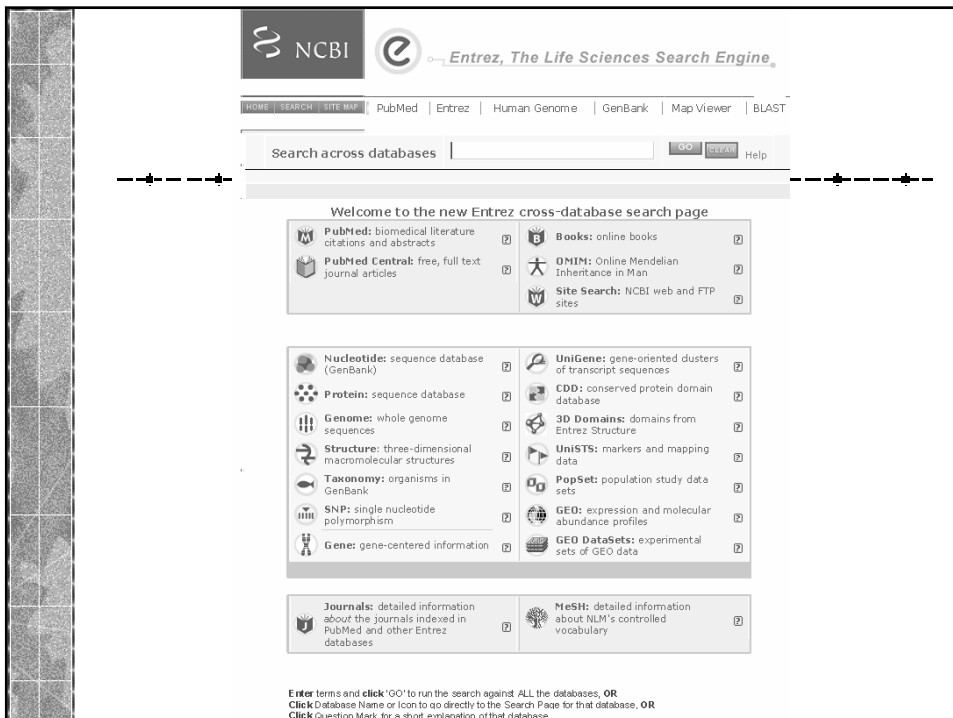
National Center for Biotechnology Information

✧ Established 1988

✧ Creates

- ◆ Public databases
- ◆ Conducts research in computational biology
- ◆ Develops software
- ◆ Disseminates biomedical information
- ◆ Entrez
 - Search and retrieval from many databases

<http://www.ncbi.nlm.nih.gov/Entrez/>



The screenshot shows the NCBI Entrez homepage. At the top, there are logos for NCBI and Entrez, with the tagline "Entrez, The Life Sciences Search Engine". Below the logos is a navigation bar with links: HOME, SEARCH, SITE MAP, PubMed, Entrez, Human Genome, GenBank, Map Viewer, and BLAST. A search bar is prominently displayed with the text "Search across databases" and buttons for GO, REPAIR, and Help. Below the search bar, a welcome message reads "Welcome to the new Entrez cross-database search page". The main content area is organized into a grid of database categories, each with an icon, a brief description, and a question mark icon for more information. The categories include: PubMed (biomedical literature citations and abstracts), PubMed Central (free, full text journal articles), Books (online books), OMIM (Online Mendelian Inheritance in Man), Site Search (NCBI web and FTP sites), Nucleotide (sequence database (GenBank)), UniGene (gene-oriented clusters of transcript sequences), Protein (sequence database), CDD (conserved protein domain database), Genome (whole genome sequences), 3D Domains (domains from Entrez Structure), Structure (three-dimensional macromolecular structures), UniSTS (markers and mapping data), Taxonomy (organisms in GenBank), PopSet (population study data sets), SNP (single nucleotide polymorphism), GEO (expression and molecular abundance profiles), Gene (gene-centered information), GEO Datasets (experimental sets of GEO data), Journals (detailed information about the journals indexed in PubMed and other Entrez databases), and MeSH (detailed information about NLM's controlled vocabulary). At the bottom, a footer provides instructions: "Enter terms and click 'GO' to run the search against ALL the databases, OR Click Database Name or icon to go directly to the Search Page for that database, OR Click Question Mark for a short explanation of that database."

NCBI Entrez, The Life Sciences Search Engine

HOME SEARCH SITE MAP PubMed Entrez Human Genome GenBank Map Viewer BLAST

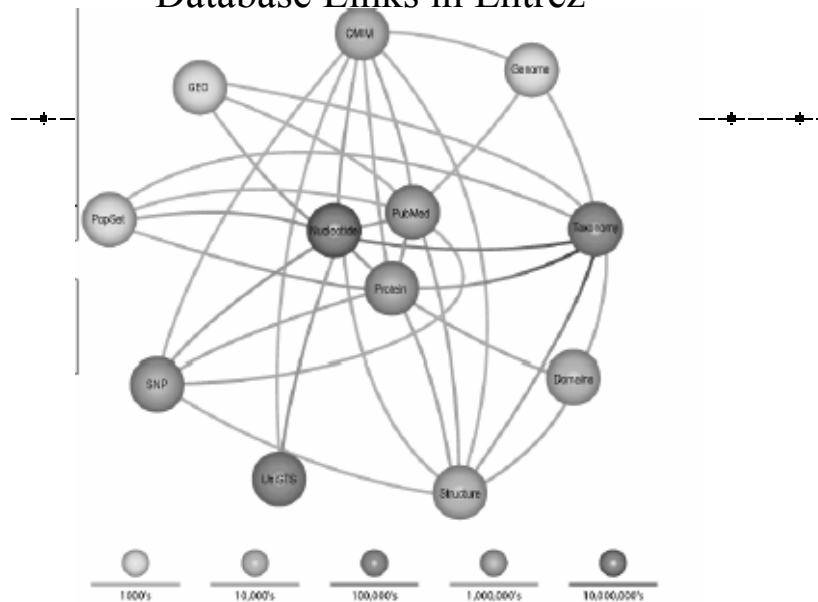
Search across databases [Help](#)

Welcome to the new Entrez cross-database search page

PubMed: biomedical literature citations and abstracts	Books: online books
PubMed Central: free, full text journal articles	OMIM: Online Mendelian Inheritance in Man
	Site Search: NCBI web and FTP sites
Nucleotide: sequence database (GenBank)	UniGene: gene-oriented clusters of transcript sequences
Protein: sequence database	CDD: conserved protein domain database
Genome: whole genome sequences	3D Domains: domains from Entrez Structure
Structure: three-dimensional macromolecular structures	UniSTS: markers and mapping data
Taxonomy: organisms in GenBank	PopSet: population study data sets
SNP: single nucleotide polymorphism	GEO: expression and molecular abundance profiles
Gene: gene-centered information	GEO Datasets: experimental sets of GEO data
Journals: detailed information about the journals indexed in PubMed and other Entrez databases	MeSH: detailed information about NLM's controlled vocabulary

Enter terms and click "GO" to run the search against ALL the databases, OR
Click Database Name or icon to go directly to the Search Page for that database, OR
Click Question Mark for a short explanation of that database

Database Links in Entrez



Jim Ostell, The NCBI Handbook, Chapter 14

Entrez: Search and Retrieval System for NCBI Databases

Organism	Phylogenetic and taxonomic information
Nucleotide	GenBank, EMBL, DDBJ, RefSeq, PDB
PopSet	Population study datasets
Genome	Complete genomes
SNP	Single nucleotide polymorphism
EST	Expressed sequence tags
UniGene	Clusters of expressed sequences
GEO	Microarray datasets
Protein	Translations of GenBank & RefSeq records, SWISS-PROT, PIR, PRF, PDB
Domains	CDD: conserved domain database
Structure	MMDB: experimental 3D structures
Pubmed	Biomedical literature
PubMed Central	Free online journals
Books	Free online textbooks

Online Books



MOLECULAR BIOLOGY OF THE CELL

NCBI

FORMER NCBI HOMEPAGE MOLECULAR BIOLOGY OF THE CELL HOMEPAGE

MOLECULAR CELL BIOLOGY

Lodish Berk Zipursky Matsudaira Baltimore Darnell

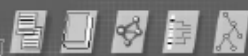
W H FREEMAN AND COMPANY

MEDICAL MICROBIOLOGY

4th Edition

Edited by Samuel Baron

The University of Texas Medical Branch at Galveston



The NCBI Handbook

NCBI

Entrez: Search and Retrieval System

<http://www.ncbi.nlm.nih.gov/Entrez/>

NCBI

Entrez

search and retrieval system

PubMed Entrez BLAST OMIM Book Taxonomy Sra data

Search PubMed for Go Clear

About Entrez

SITE MAP

PubMed Help
help documentation
for PubMed

Entrez Help
help documentation

Entrez Tutorial
tutorial for the Entrez
system

The Entrez Databases
cross-references and
information



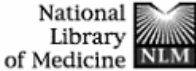
Batch Entrez
upload a file of GI or
accession numbers
to retrieve sequences

Making WWW
Links to Entrez
linking to PubMed

Entrez is a retrieval system for searching several linked databases. It provides access to:

PubMed: biomedical literature
Nucleotide: sequence database (GenBank)
Protein: sequence database
Structure: three-dimensional macromolecular structures
Genome: complete genome assemblies
Books: BookShelf online books
Domains: conserved domains (CDD)
3D Domains: domains from Entrez Structure
GEO: Gene Expression Omnibus
GEO Datasets: curated GEO data sets
Journals: journals in Entrez
MeSH: medical subject headings
NCBI Web Site: NCBI Web site search
OMIM: Online Mendelian Inheritance in Man
PMC: full-text digital archive of life sciences journal literature
PopSet: population study datasets
SNP: single nucleotide polymorphisms
Taxonomy: organisms in GenBank
UniGene: gene-oriented clusters of transcript sequences
UniSTS: markers and mapping data

NCBI

[Entrez](#)
[PubMed](#)
[Nucleotide](#)
[Protein](#)
[Genome](#)
[Structure](#)
[Databases](#)
[PMC](#)
[Journals](#)
[Books](#)

Search for



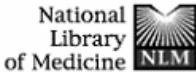
[Limits](#)
[Preview/Index](#)
[History](#)
[Clipboard](#)
[Details](#)

Entrez PubMed

- Enter one or more search terms, or click [Preview/Index](#) for advanced searching.
- Enter author names as smith jc. Initials are optional.
- Enter journal titles in full or as MEDLINE abbreviations. Use the Journals Database to find journal titles.

PubMed Services

PubMed, a service of the National Library of Medicine, includes over 14 million citations for biomedical articles back to the 1950's. These citations are from MEDLINE and additional life science journals. PubMed includes links to many sites providing full text articles and abstracts.

[Entrez](#)
[PubMed](#)
[Nucleotide](#)
[Protein](#)
[Genome](#)
[Structure](#)
[Databases](#)
[PMC](#)
[Journals](#)
[Books](#)

Search for

[Limits](#)
[Preview/Index](#)
[History](#)
[Clipboard](#)
[Details](#)

Entrez PubMed

- Enter one or more search terms, or click [Preview/Index](#) for advanced searching.
- Enter author names as smith jc. Initials are optional.
- Enter journal titles in full or as MEDLINE abbreviations. Use the Journals Database to find journal titles.

PubMed Services

PubMed, a service of the National Library of Medicine, includes over 14 million citations for biomedical articles back to the 1950's. These citations are from MEDLINE and additional life science journals. PubMed includes links to many sites providing full text articles and abstracts.

[Entrez](#)
[PubMed](#)
[Nucleotide](#)
[Protein](#)
[Genome](#)
[Structure](#)
[OMIM](#)
[PMC](#)
[Journals](#)
[Books](#)

Search

PubMed

for soybean

Go

Limits

Preview/Index

History

Clipboard

Details

- Use All Fields pull-down menu to specify a field.
- Boolean operators AND, OR, NOT must be in upper case.
- If search fields tags are used enclose in square brackets, e.g., rubella[ti]
- Search limits may exclude in process and publisher supplied citations.

Limited to:

All Fields

only items with abstracts

Publication Types

Languages

Subsets

Ages

Human or Animal

Gender

Entrez Date

Publication Date

From

To


Use the format YYYY/MM/DD; month and day are optional.


All Fields

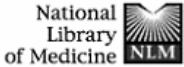
✧ Author

✧ Journal

✧ Issue







Entrez
PubMed
Nucleotide
Protein
Genome
Structure
OMIM
PMC
Journals
Books

Search

PubMed

for

soybean

Go

Limits
Preview/Index
History
Clipboard
Details

Enter terms and click Preview to see only the number of search results.

To combine searches use # before search number, e.g., (#2 OR #3) AND asthma.

No history available

Add Term(s) to Query or View Index:


- Enter a term in the text box; use the pull-down menu to specify a search field.
- Click Preview to add terms to the query box and see the number of search results, or click Index to view terms within a field.


➡

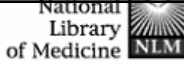
All Fields

Chalcone

Click to add a term to the query box.







Entrez
PubMed
Nucleotide
Protein
Genome
Structure
OMIM
PMC
Journals
Books

Search

PubMed

for

Soybean AND nematode

Go

Limits
Preview/Index
History
Clipboard
Details

- Use All Fields pull-down menu to specify a field.
- Boolean operators AND, OR, NOT must be in upper case. ➡
- If search fields tags are used enclose in square brackets, e.g., rubella [ti]
- Search limits may exclude in process and publisher supplied citations.

Limited to:

All Fields

▾

☐ only items with abstracts

Publication Types ▾

Languages ▾

Subsets ▾

Ages ▾

Human or Animal ▾

Gender ▾

Entrez Date ▾

Publication Date ▾

From

To

Use the format YYYY/MM/DD; month and day are optional.

Keyword searches

- ✧ Soybean
- ✧ Soybean nematode chalcone
- ✧ Soybean AND nematode NOT chalcone
 - Boolean operators: “AND” “OR” “NOT”
- ✧ “soybean cyst nematode”

The screenshot shows the PubMed website interface. At the top, there are logos for NCBI, PubMed, and the National Library of Medicine (NLM). Below these are navigation links: Entrez, PubMed, Nucleotide, Protein, Genome, Structure, OMIM, PMC, Journals, and Books. A search bar is present with a dropdown menu set to 'PubMed' and a 'Go' button. Below the search bar are tabs for Limits, Preview/Index, History (which is circled), Clipboard, and Details. A mouse cursor is pointing at the History tab. Below the tabs, there is a list of search history items with a bulleted list of instructions: Search History will be lost after eight hours of inactivity; To combine searches use # before search number, e.g., #2 AND #6; Search numbers may not be continuous; all searches are represented. Below this list is a table titled 'Most Recent Queries' with columns for Search, Most Recent Queries, Time, and Result. The table lists three searches: #3 Search soybean and nematode AND chalcone (Time: 15:56:50, Result: 2), #2 Search soybean and nematode (Time: 15:56:24, Result: 83), and #1 Search soybean Field: All Fields (Time: 15:56:05, Result: 17767). On the left side of the page, there are links for Entrez PubMed, PubMed Services, and Related Resources.

NCBI PubMed National Library of Medicine NLM

Entrez PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search PubMed for Go

Limits Preview/Index **History** Clipboard Details

- Search History will be lost after eight hours of inactivity.
- To combine searches use # before search number, e.g., #2 AND #6.
- Search numbers may not be continuous; all searches are represented.

Search	Most Recent Queries	Time	Result
#3	Search soybean and nematode AND chalcone	15:56:50	2
#2	Search soybean and nematode	15:56:24	83
#1	Search soybean Field: All Fields	15:56:05	17767

Entrez PubMed

PubMed Services

Related Resources

[Entrez](#)
[PubMed](#)
[Nucleotide](#)
[Protein](#)
[Genome](#)
[Structure](#)
[OMM](#)
[PMC](#)
[Journals](#)
[Books](#)

Search

PubMed

For

chalcone soybean

Go

[Limits](#)
[Preview/Index](#)
[History](#)
[Clipboard](#)
[Details](#)

Summary

Show: 20

Sort

Text

Items 1-20 of 48

1 of 3

Next

1: [Lewers K, Heine R, Beard H, Mark L, Matthews B.](#)

Related Articles, Links

A physical map of a gene-dense region in soybean linkage group A2 near the black seed coat and Rhg (4) loci. *Theor Appl Genet.* 2002 Feb;104(2-3):254-260. PMID: 12382695 [PubMed - as supplied by publisher]

2: [Senda M, Yumoto A, Yumoto S, Ishikawa R, Harada T, Niizeki M, Akada S.](#)

Related Articles, Links

Analysis of the duplicated CHS1 gene related to the suppression of the seed coat pigmentation in yellow soybeans. *Theor Appl Genet.* 2002 May;104(6-7):1086-1091. Epub 2002 Feb 20. PMID: 12382616 [PubMed - as supplied by publisher]

3: [Senda M, Kase A, Yumoto S, Akada S, Ishikawa R, Harada T, Niizeki M.](#)

Related Articles, Links

Sequence divergence at chalcone synthase gene in pigmented seed coat soybean mutants of the Inhibitor locus. *Genes Genet Syst.* 2002 Oct;77(5):341-50. PMID: 12441645 [PubMed - indexed for MEDLINE]

4: [Liu CJ, Blount JW, Steele CL, Dixon RA.](#)

Related Articles, Links

Bottlenecks for metabolic engineering of isoflavone glycoconjugates in *Arabidopsis*. *Proc Natl Acad Sci U S A.* 2002 Oct 29;99(22):14578-83. Epub 2002 Oct 16. PMID: 12384577 [PubMed - indexed for MEDLINE]

[Entrez](#)
[PubMed](#)
[Nucleotide](#)
[Protein](#)
[Genome](#)
[Structure](#)
[OMM](#)
[PMC](#)
[Journals](#)
[Books](#)

Search

PubMed

For

chalcone soybean

Go

[Limits](#)
[Preview/Index](#)
[History](#)
[Clipboard](#)
[Details](#)

Summary

Show: 20

Sort

Text

Items 1-20 of 48

1 of 3

Next

1: [Lewers K, Heine R, Beard H, Mark L, Matthews B.](#)

Related Articles, Links

A physical map of a gene-dense region in soybean linkage group A2 near the black seed coat and Rhg (4) loci. *Theor Appl Genet.* 2002 Feb;104(2-3):254-260. PMID: 12382695 [PubMed - as supplied by publisher]

2: [Senda M, Yumoto A, Yumoto S, Ishikawa R, Harada T, Niizeki M, Akada S.](#)

Related Articles, Links

Analysis of the duplicated CHS1 gene related to the suppression of the seed coat pigmentation in yellow soybeans. *Theor Appl Genet.* 2002 May;104(6-7):1086-1091. Epub 2002 Feb 20. PMID: 12382616 [PubMed - as supplied by publisher]

3: [Senda M, Kase A, Yumoto S, Akada S, Ishikawa R, Harada T, Niizeki M.](#)

Related Articles, Links

Sequence divergence at chalcone synthase gene in pigmented seed coat soybean mutants of the Inhibitor locus. *Genes Genet Syst.* 2002 Oct;77(5):341-50. PMID: 12441645 [PubMed - indexed for MEDLINE]

4: [Liu CJ, Blount JW, Steele CL, Dixon RA.](#)

Related Articles, Links

Bottlenecks for metabolic engineering of isoflavone glycoconjugates in *Arabidopsis*. *Proc Natl Acad Sci U S A.* 2002 Oct 29;99(22):14578-83. Epub 2002 Oct 16. PMID: 12384577 [PubMed - indexed for MEDLINE]

[Entrez](#)
[PubMed](#)
[Nucleotide](#)
[Protein](#)
[Genome](#)
[Structure](#)
[OMIM](#)
[PMC](#)
[Journals](#)
[Books](#)

Search
for

[Limits](#)
[Preview/Index](#)
[History](#)
[Clipboard](#)
[Details](#)

Summary
Show:
Sort:

Items 1-2 of 2
One page.

☐
1:
[Lewers K, Heinz R, Beard H, Marek L, Matthews B.](#)
[Related Articles, Links](#)

A physical map of a gene-dense region in soybean linkage group A2 near the black seed coat and Rhg (4) loci. Theor Appl Genet. 2002 Feb;104(2-3):254-260. PMID: 12582695 [PubMed - as supplied by publisher]

☐
2:
[Senda M, Kasai A, Yanoto S, Akada S, Ishikawa R, Harada T, Niizeki M.](#)
[Related Articles, Links](#)

Sequence divergence at chalcone synthase gene in pigmented seed coat soybean mutants of the Inhibitor locus. Genes Genet Syst. 2002 Oct;77(5):341-50. PMID: 12441645 [PubMed - indexed for MEDLINE]

[Entrez](#)
[PubMed](#)
[Nucleotide](#)
[Protein](#)
[Genome](#)
[Structure](#)
[OMIM](#)
[PMC](#)
[Journals](#)
[Books](#)

Search
for

[Limits](#)
[Preview/Index](#)
[History](#)
[Clipboard](#)
[Details](#)

Abstract
Show:
Sort:

☐
1: Theor Appl Genet. 2002 Feb;104(2-3):254-260.
[Related Articles, Links](#)

A physical map of a gene-dense region in soybean linkage group A2 near the black seed coat and Rhg (4) loci.

Lewers K, Heinz R, Beard H, Marek L, Matthews B.

USDA ARS PSI, Soybean Genomics and Improvement Laboratory, Bldg 006 BARC-West, Beltsville, MD 20705, USA. matthewb@ba.ars.usda.gov

Soybean (Glycine max L. Merrill) linkage group A2 contains a major resistance gene to the soybean cyst nematode (Heterodera glycines Ichinohe) at the Rhg (4) locus near a gene encoding aspartokinase homoserine dehydrogenase (AK-HSDH) and also near the I locus affecting seed coat color. To

NCBI PubMed National Library of Medicine NLM

Entrez PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search PubMed for Go

Limits Preview/Index History Clipboard Details

Entrez PubMed

Summary Show: 20 Sort Text

Items 1-20 of 105

1 of 6 Next

PubMed Services

1: [Lewers K, Heirz R, Beard H, Marek L, Matthews B.](#) Related Articles, Links

A physical map of a gene-dense region in soybean linkage group A2 near the black seed coat and Rhg (4) loci. Theor Appl Genet. 2002 Feb;104(2-3):254-260. PMID: 12582695 [PubMed - as supplied by publisher]

Related Resources

2: [Lewers KS, Nimalgoda SD, Warner AL, Knap HT, Matthews BF.](#) Related Articles, Links

Physical mapping of resistant and susceptible soybean genomes near the soybean cyst nematode resistance gene Rhg4. Genome. 2001 Dec;44(6):1057-64. PMID: 11768209 [PubMed - indexed for MEDLINE]

Other searches

✳ Author's name-query

- ◆ Find other papers by the author that may be related to the topic
- ◆ Find paper you remember author but not source

✳ Field-restricted query

- ◆ Narrows search
- ◆ Dates, address, journal

Limits

- ✧ Narrow the focus of a search
- ✧ Publication types

The screenshot shows the PubMed website in a Microsoft Internet Explorer browser window. The address bar displays <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi>. The page features the NCBI logo on the left and the PubMed logo with the National Library of Medicine (NLM) logo on the right. A navigation bar includes links for Entrez, PubMed, Nucleotide, Protein, Genome, Structure, OMIM, PMC, Journals, and Books. Below this, a search bar is present with 'Limits', 'Preview/Index', 'History', 'Clipboard', and 'Details' tabs. The 'Limits' tab is currently selected and circled. The main content area contains instructions for using search limits, such as entering search terms, author names, or journal titles. It also includes sections for 'Bookshelf Additions' (listing books like 'Cancer Medicine, 6th ed.' and 'The Genetic Landscape of Diabetes') and a 'New Clinical Alert' (dated March 2, 2004, regarding the Women's Health Initiative Estrogen-Alone Study). A sidebar on the left provides links to various PubMed services and related resources. The footer mentions the 'New Global NCBI Search Engine'.

Entrez PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books


Search PubMed

Limits Preview/Index History Clipboard Details

- Enter one or more search terms, or click [Preview/Index](#) for advanced searching.
- Enter author names as smith jc. Initials are optional.
- Enter journal titles in full or as MEDLINE abbreviations. Use the [Journals Database](#) to find journal titles.

PubMed, a service of the National Library of Medicine, includes over 14 million citations for biomedical articles back to the 1950's. These citations are from MEDLINE and additional life science journals. PubMed includes links to many sites providing full text articles and other related resources.

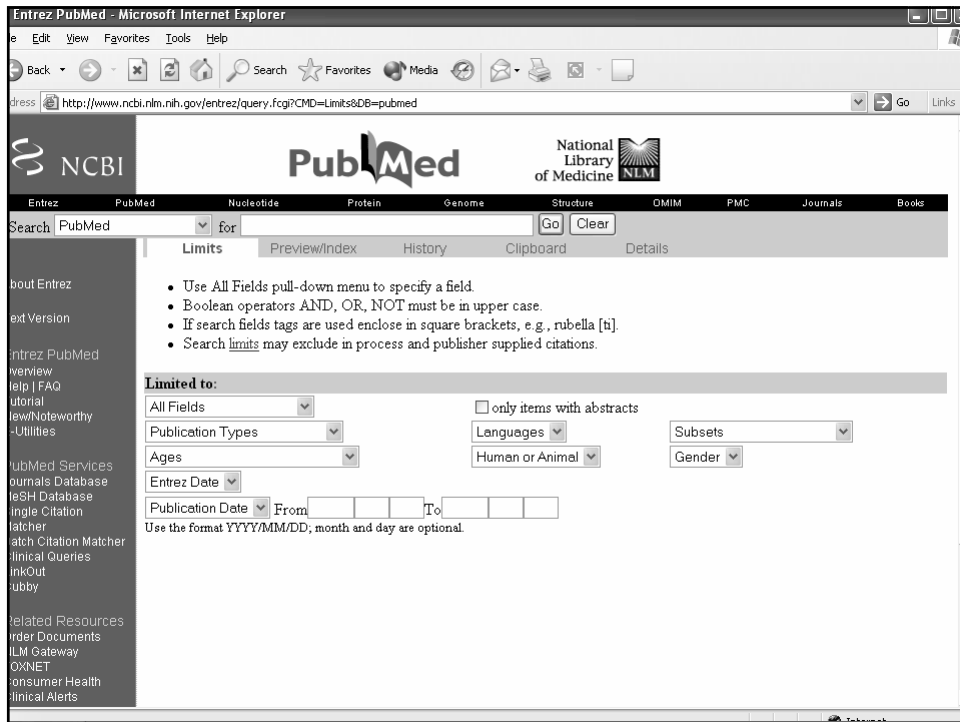
Bookshelf Additions

 *Cancer Medicine, 6th ed. and The Genetic Landscape of Diabetes* are now available for interactive searching on the [Bookshelf](#).

New Clinical Alert

On March 2, 2004 the National Heart Lung and Blood Institute (NHLBI) has issued a new [clinical alert](#): NIH Asks Participants in Women's Health Initiative Estrogen-Alone Study to Stop Study Pills, Begin Follow-up Phase.

New Global NCBI Search Engine



Entrez PubMed - Microsoft Internet Explorer

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=Limits&DB=pubmed

NCBI PubMed National Library of Medicine NLM

Entrez PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search PubMed Limits Preview/Index History Clipboard Details

about Entrez
next Version
Entrez PubMed
view
help | FAQ
tutorial
new/Noteworthy
Utilities
PubMed Services
Journals Database
eSH Database
Single Citation
atcher
atch Citation Matcher
Clinical Queries
LinkOut
ubby
Related Resources
Order Documents
ILM Gateway
OXNET
Consumer Health
Clinical Alerts

- Use All Fields pull-down menu to specify a field.
- Boolean operators AND, OR, NOT must be in upper case.
- If search fields tags are used enclose in square brackets, e.g., rubella [ti].
- Search limits may exclude in process and publisher supplied citations.

Limited to:

All Fields ☐ only items with abstracts

Publication Types ☐ Languages Subsets

Clinical Trial ☐ Human or Animal Gender

Editorial

Letter

Meta-Analysis

Practice Guideline

Randomized Controlled Trial

Review

From To

h and day are optional.

Entrez PubMed - Microsoft Internet Explorer

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=Limits&DB=pubmed

NCBI PubMed National Library of Medicine NLM

Entrez PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search PubMed Limits Preview/Index History Clipboard Details

about Entrez
next Version
Entrez PubMed
view
help | FAQ
tutorial
new/Noteworthy
Utilities
PubMed Services
Journals Database
eSH Database
Single Citation
atcher
atch Citation Matcher
Clinical Queries
LinkOut
ubby
Related Resources
Order Documents
ILM Gateway
OXNET
Consumer Health
Clinical Alerts

- Use All Fields pull-down menu to specify a field.
- Boolean operators AND, OR, NOT must be in upper case.
- If search fields tags are used enclose in square brackets, e.g., rubella [ti].
- Search limits may exclude in process and publisher supplied citations.

Limited to:

All Fields ☐ only items with abstracts

Publication Types ☐ Languages Subsets

Ages ☐ Human or Animal Gender

Entrez Date

Publication Date From To

Use the format YYYY/MM/DD; month and day are optional.

Entrez PubMed - Microsoft Internet Explorer

Address: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=Limits&DB=pubmed

NCBI PubMed National Library of Medicine NLM

Entrez PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search PubMed Limits Preview/Index History Clipboard Details

about Entrez
next Version
Entrez PubMed
view
help | FAQ
tutorial
new/Noteworthy
Utilities
PubMed Services
Journals Database
eSH Database
Single Citation
atcher
atch Citation Matcher
Clinical Queries
LinkOut
Pubby

Related Resources
Order Documents
ILM Gateway
OXNET
Consumer Health
Clinical Alerts

- Use All Fields pull-down menu to specify a field.
- Boolean operators AND, OR, NOT must be in upper case.
- If search fields tags are used enclose in square brackets, e.g., rubella [ti].
- Search limits may exclude in process and publisher supplied citations.

Limited to:

All Fields ☐ only items with abstracts

Publication Types Languages Subsets

Ages Human or Animal Gender

Entrez Date

Publication Date From 1998 To 2004

Use the format YYYY/MM/DD; month and day are optional

Entrez PubMed - Microsoft Internet Explorer

Address: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=Limits&DB=pubmed

NCBI PubMed National Library of Medicine NLM

Entrez PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search PubMed Limits Preview/Index History Clipboard Details

about Entrez
next Version
Entrez PubMed
view
help | FAQ
tutorial
new/Noteworthy
Utilities
PubMed Services
Journals Database
eSH Database
Single Citation
atcher
atch Citation Matcher
Clinical Queries
LinkOut
Pubby

Related Resources
Order Documents
ILM Gateway
OXNET
Consumer Health
Clinical Alerts

- Use All Fields pull-down menu to specify a field.
- Boolean operators AND, OR, NOT must be in upper case.
- If search fields tags are used enclose in square brackets, e.g., rubella [ti].
- Search limits may exclude in process and publisher supplied citations.

Limited to:

All Fields ☐ only items with abstracts

Publication Types Languages Subsets

Ages Human or Animal Gender

Entrez Date

Publication Date From To

Use the format YYYY/MM/DD; month and day are optional

Subsets

- Subsets
- AIDS
- Bioethics
- Cancer
- Complementary Medicine
- Core clinical journals
- Dental journals
- History of Medicine
- MEDLINE
- Nursing journals
- OLDMEDLINE for Pre1966
- PubMed Central
- Space Life Sciences
- Toxicology

Tips

- ✖ Quoted queries behave as a single word
 - ◆ Example ... “Soybean cyst nematode”
- ✖ Add initials to author queries
 - ◆ Example... Matthews B
- ✖ If you don't find what you want, don't give up
 - ◆ Try a different key word
 - ◆ Indexing terminology is important
- ✖ Try “Related papers” button
- ✖ No papers before 1965 are here
- ✖ No abstracts for papers before 1976 are here

Agricola

- ✖ **AGRICultural On Line Access**
- ✖ Bibliographic database of citations to agricultural literature
- ✖ Two databases
 - ◆ NAL catalog
 - Books, serials, audio-visuals
 - ◆ Article Citation Index
 - Journal articles
- ✖ U.S. Department of Agriculture
- ✖ National Agricultural Library
- ✖ <http://agricola.nal.usda.gov/>

PubMed vs Agricola

✱ Substantial overlap of journals indexed

✱ PubMed

- ◆ Broader molecular, medicine, biology
- ◆ Extensive
- ◆ No basic agronomy

✱ Agricola

- ◆ Agriculture-related
- ◆ Narrower focus
- ◆ Basic agronomy found here

[Accessible Site](#)

National Agricultural Library Article Citation Database **AGRICOLA**

NAL Catalog

Includes books, serials, audiovisuals, and other resources.

Quick Search: (enter single word or phrase)

More search options:

[Search ...](#) by keyword using forms and menus
[Browse ...](#) by author, title, or subject

Article Citation Database

Includes journal articles, book chapters, short reports, and reprints. See [Journals Indexed in AGRICOLA](#)

Quick Search: (enter single word or phrase)

More search options:

[Search ...](#) by keyword using forms and menus
[Browse ...](#) by author, title, or subject

How to Search Both Databases at Once

Some search options are not available when you search both databases simultaneously.

[firstgov.gov](#) | [science.gov](#) | [USDA](#) | [ARS](#) | [NAL](#) | [What is AGRICOLA?](#) | [Help](#) |

Comments or questions about this database may be sent to AGRICOLAhelp@nal.usda.gov.

National Agricultural Library **AGRICOLA** Article Citation Database

[Databases](#) [New Search](#) [Request an Item](#) [Help](#) [Home](#)

Database Name: Article Citation Database

Search	Browse	
Search: <input type="text"/>	all of these <input type="button" value="v"/>	In: <input type="text" value="Keyword Anywhere"/>
<input type="radio"/> AND <input type="radio"/> OR <input type="radio"/> NOT		
Search: <input type="text"/>	any of these <input type="button" value="v"/>	In: <input type="text" value="Subject"/>
<input type="radio"/> AND <input type="radio"/> OR <input type="radio"/> NOT		
Search: <input type="text"/>	as a phrase <input type="button" value="v"/>	In: <input type="text" value="Title"/>
<input type="radio"/> AND <input type="radio"/> OR <input type="radio"/> NOT		
Search: <input type="text"/>	as a phrase <input type="button" value="v"/>	In: <input type="text" value="Author Name"/>
<input type="radio"/> AND <input type="radio"/> OR <input type="radio"/> NOT		
25 records per page <input type="button" value="v"/>		

National Agricultural Library **AGRICOLA** Article Citation Database

[Databases](#) [New Search](#) [Request an Item](#) [Help](#) [Home](#)

Database Name: Article Citation Database

Search	Browse	
Search: <input type="text" value="soybean"/>	all of these <input type="button" value="v"/>	In: <input type="text" value="Keyword Anywhere"/>
<input type="radio"/> AND <input type="radio"/> OR <input type="radio"/> NOT		
Search: <input type="text" value="nematode"/>	any of these <input type="button" value="v"/>	In: <input type="text" value="Keyword Anywhere"/>
<input type="radio"/> AND <input type="radio"/> OR <input type="radio"/> NOT		
Search: <input type="text" value="chalcone"/>	as a phrase <input type="button" value="v"/>	In: <input type="text" value="Keyword Anywhere"/>
<input type="radio"/> AND <input type="radio"/> OR <input type="radio"/> NOT		
Search: <input type="text"/>	as a phrase <input type="button" value="v"/>	In: <input type="text" value="Author Name"/>
<input type="radio"/> AND <input type="radio"/> OR <input type="radio"/> NOT		
25 records per page <input type="button" value="v"/>		
<input type="button" value="Search"/> <input type="button" value="Reset"/>		<input type="button" value="Set Limits"/>

[Search Hints](#)

National Agricultural Library **AGRICO** Article Citation Database

[Databases](#) [New Search](#) [Titles List](#) [Request an Item](#) [Search History](#) [Help](#) [Home](#)

Database Name: Article Citation Database

Search Request: Search = (soybean)[in Keyword Anywhere] AND (nematode)[in Keyword Anywhere] AND ("chalcone")[in Keyword Anywhere]

Search Results: Displaying 1 through 2 of 2 entries.

#	Date	Title Long	Author
<input type="checkbox"/> [1]	2002	Physical map of a gene-dense region in soybean linkage group A2 near the black seed coat and Rhg4 loci.	Lewers, K.
<input type="checkbox"/> [2]	1999	Expression of soybean cyst nematode resistance in transgenic hairy roots of soybean.	Narayanan, R. A.

[Clear Selected](#) [Retain Selected](#)

Print, Save or Email	
Select Records <input type="radio"/> All on this page <input type="radio"/> Marked on this page <input type="radio"/> Marked on all pages	Full Record <input type="button" value="Format for Print or Save"/> Select Format: <input type="text"/> Enter your email address: <input type="text"/>

What we covered today

- ✱ What is bioinformatics
- ✱ Databases
- ✱ Literature searches